

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2001-522471

(P2001-522471A)

(43) 公表日 平成13年11月13日 (2001. 11. 13)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 1 0 L 21/04		G 1 0 K 15/04	3 0 2 D
G 1 0 K 15/04	3 0 2	G 1 0 L 3/02	A

審査請求 未請求 予備審査請求 有 (全 39 頁)

(21) 出願番号 特願平10-546443
 (86) (22) 出願日 平成10年4月27日 (1998. 4. 27)
 (85) 翻訳文提出日 平成11年10月28日 (1999. 10. 28)
 (86) 国際出願番号 PCT/CA98/00406
 (87) 国際公開番号 WO98/49670
 (87) 国際公開日 平成10年11月5日 (1998. 11. 5)
 (31) 優先権主張番号 08/848, 050
 (32) 優先日 平成9年4月28日 (1997. 4. 28)
 (33) 優先権主張国 米国 (US)

(71) 出願人 アイブイエル テクノロジーズ エルティ
 ーディー。
 カナダ国 プリティッシュ コロンビア
 ブイ8エム 1セツト6, ビクトリア, パ
 ートラン プレイス 6710
 (72) 発明者 ギブソン, ブライアン チャールズ
 カナダ国 プリティッシュ コロンビア
 ブイ8エックス 3エックスアイ, ビクト
 リア, ヘクター ロード 235
 (74) 代理人 弁理士 廣江 武典

最終頁に続く

(54) 【発明の名称】 特定の声を目標とする音声変換

(57) 【要約】

本発明は、目標たる個人の声の特性に合わせるため、音声源個人の声を変換する方法である。前記の目標個人の声の励起信号成分を抽出して、音声源個人の声のスペクトルエンベロープも抽出する。変換した音声、前記の音声源個人のスペクトルエンベロープを目標個人の声の励起信号成分に適用することにより分析する。そして、前記信号の無声区域や不適切な発声区域を適切な発声区域からのデータで置き換えて作成した改良励起信号を使えば、高い品質の変換が達成できる。また、音声源個人の声のスペクトル特性を変換する方法も、いくつか開示している。

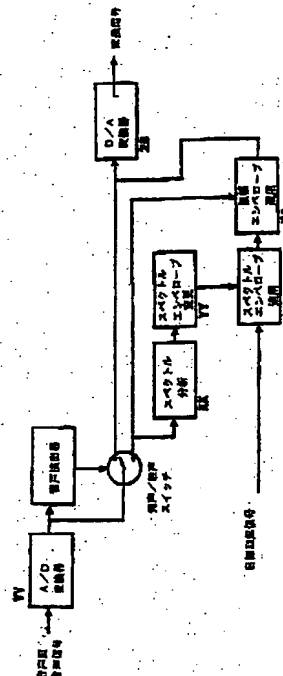


図3

【特許請求の範囲】

1. 目標たる個人の特徴に合わせるため発声源個人の声を変換する方法であって、

前記の目標たる個人の声から由来するスペクトルエンベロープを、目標たる個人の声からする励起信号成分に適用する工程から成る方法。

2. さらに、前記の目標たる個人の声から前記の励起信号成分を抽出する工程とから成る、請求項1記載の変換方法。

3. さらに、前記の励起信号成分を保存する工程と、
前記の音声信号のスペクトルエンベロープを決めるため、発声源個人の声の個別音声信号のスペクトル分析を実行する工程とから成る、請求項1記載の変換方法。

4. さらに、前記の目標たる個人の個別音声信号のピッチを決める工程とから成る、請求項1記載の変換方法。

5. さらに、前記の発声源の音声信号のピッチと一致するよう、目標たる励起信号のピッチを変換する工程とから成る、請求項4記載の変換方法。

6. 前記の励起信号を抽出する工程が、目標たる音声信号のスペクトルエンベロープを平坦化することにより実行できる、請求項2記載の変換方法。

7. さらに、前記の発声源個人の声の個別信号を発声区域と無声区域に分割する工程と、

所定の区域が発声入力を示す場合には、前記の区域から由来するスペクトルエンベロープを前記の励起信号成分に適用して出力を生成する工程と、

所定の区域が無声入力を示す場合には、前記の励起信号成分を参照することなく、前記区域に基づいた出力を生成する工程とから成る、請求項1記載の変換方法。

8. さらに、前記の励起信号成分を保存する工程とから成る、請求項2記載の変換方法。

9. 前記の保存する工程が、前記の励起信号成分を圧縮形式で保存する工程である、請求項2記載の変換方法。

10. さらに、前記の音声信号のスペクトルエンベロープを前記の励起信号に適

用するより前に、前記の音声信号をスペクトルエンベロープを変換する工程とから成る、請求項1または3記載の変換方法。

11. 前記の発声源個人の声から由来するスペクトルエンベロープを適用する工程が、前記の信号を複数の周波数帯に分割する工程と、前記スペクトルエンベロープを前記周波数帯に対応させて個別に変換する工程と、前記の変換されたスペクトルエンベロープを前記周波数帯に適用する工程とから成る、請求項1記載の変換方法。

12. 目標たる個人の特徴に合わせるため発声源個人の声を変換する方法であって、

その目標たる個人の声の個別音声信号を保存する工程と、

前記の音声信号の励起信号成分を抽出する工程とから成る方法。

13. さらに、前記の抽出した励起信号を保存する工程とから成る、請求項12記載の変換方法。

14. さらに、前記の音声信号のスペクトルエンベロープを前記の励起信号に適用するよりも前に、前記の第2の音声信号のスペクトルエンベロープを変換する工程とから成り、その変換する工程が、スペクトル分析を実行する工程よりも前に、音声源個人の声の個別音声信号のサンプルの1ブロック分の時間伸長を変更する工程から成る、請求項1記載の変換方法。

15. 前記の音声源個人と目標たる個人が歌手である、請求項1から14のいずれか記載の変換方法。

16. 前記の励起信号を抽出する工程が、

その時間変動スペクトルエンベロープを決めるため、目標たる音声信号のスペクトル分析を実行する工程と、

時間変動フィルターを作成するため前記のスペクトルエンベロープを利用する工程と、

前記スペクトルエンベロープを平坦化するため前記の時間変動フィルターを利用する工程とから成る、請求項2記載の変換方法。

17. さらに、発声セグメントと無声セグメントとを特定し、その無声セグメントを音声データと差し替える工程とから成る、請求項16記載の変換方法。

18. 前記の信号中の無声セグメントが、そのセグメントのパラメータを、平均セグメント出力値、平均低帯域セグメント出力値、セグメントに対するゼロ交差数から成るパラメータグループから選択された閾値と比較することにより特定できる、請求項17記載の変換方法。

19. 前記の発生データで置き換える工程が、無声部分に隣接した音声信号の端部間でモーフィング処理できるよう正弦波合成を利用する工程から成る、請求項17記載の変換方法。

20. 信号の2つの音声部分間を補間する方法であって、信号のピッチ輪郭線を算定する工程と、音声部分の端部でのスペクトルの振幅と位相を得るため、ピーク選別にてスペクトル分析を行う工程と、境界部での位相連続性を確保できるよう、線形周波数補正項を含み、補間されたピッチ輪郭線により限定される正弦波合成方法を使う工程とから成る補間方法。

21. さらに、ランダムピッチ成分を利用する工程とから成る、請求項20記載の補間方法。

22. 音声信号から励起信号を抽出する方法であって、前記の音声信号のセグメントが発声信号であるか無声信号であるかを判断する工程と、

前記の発声信号を示すセグメントのピッチを算定して保存する工程と、その時間変動スペクトルエンベロープを算定できるよう、前記の音声信号のスペクトル分析を行う工程と、

時間変動フィルターを作成するため前記のスペクトルエンベロープを使う工程と、

前記のスペクトルエンベロープを平坦化するため前記の時間変動フィルターを使う工程とから成る抽出方法。

23. 前記のセグメントが発声信号であるか無声信号であるかを判断する工程が、前記のセグメントのパラメータを、平均セグメント出力値、平均低帯域セグメント出力値、セグメント1個あたりのゼロ交差数のパラメータグループから選択された閾値と比較する工程とから成る、請求項22記載の抽出方法。

24. さらに、無声信号セグメントを音声データで置き換える工程とから成る、

請求項22記載の抽出方法。

25. 前記の音声データで置き換える工程が、前記の無声部分に隣接した発声部分へ適用するのに正弦波合成を使う工程から成る、請求項24記載の抽出方法。

26. 少なくとも2人の目標たる個人の声の特性に合わせるため音声源の個人の声を変換する方法であって、前記の音声源個人の声から由来するスペクトルエンベロープを、前記の目標たる個人の声から由来する合成励起信号に適用する工程から成る変換方法。

27. さらに、前記の目標たる個人の各自の声から励起信号成分を抽出して保存する工程と、

前記の目標たる個人の各自の声から抽出した励起信号を合成して合成励起信号にする工程と、

前記の音声信号のスペクトルエンベロープを算定できるよう、音声源個人の声を示す音声信号のスペクトル分析を行う工程とから成る、請求項26記載の変換方法。

28. 音声信号のスペクトルエンベロープを変換する方法であって、等角写像を時間変動合成フィルターの微分演算式に適用する工程から成る変換方法。

29. 前記の音声源個人と目標たる個人の少なくとも一方が歌手であって、さらに、請求項28記載の方法の工程とから成る、請求項3記載の変換方法。

30. 音声源個人の声を示す音声信号のスペクトルエンベロープを変換する方法であって、

前記の音声信号のスペクトルエンベロープに相当するデジタル変換関数を獲得する工程と、

前記のデジタル変換関数を複数の低階位部分に分解する工程と、
前記の低階位部分の少なくとも1つのスペクトル特性を変更する工程とから成る変換方法。

31. さらに、前記の第2信号のスペクトルエンベロープを前記の励起信号に適用するより前に、前記の第2信号のスペクトルエンベロープを変換する工程と、
前記の音声源音声信号の振幅エンベロープを算定する工程と、

前記の音声源個人の声のスペクトルエンベロープを前記の目標たる個人の声か

ら由来する励起信号に適用した結果からの出力信号に、前記の振幅エンベロープを適用する工程とから成る、請求項1記載の変換方法。

32. 前記の音声信号は歌唱を示すものである、請求項28記載の変換方法。

33. さらに、前記の音声源個人の声を示す音声信号を低周波数帯域と高周波数帯域に分割して、前記の低周波数帯域のみを請求項1記載の方法に従って処理する工程とから成る、請求項1記載の変換方法。

34. 前記の各帯域のスペクトルエンベロープを変換および適用する工程が、低効果サンプリングレートにて再サンプリング信号 $SD(t)$ を作成できるよう、前記帯域での信号をサンプリングする工程と、

$SD(T)$ の低階位スペクトル分析を行い、直接式フィルター係数 $aD(t)$ を算出する工程と、

前記のスペクトルをスケール処理できるよう、等角写像を使って係数 $aD(t)$ を変更する工程と、

前記の結果たるフィルターを前記の目標励起信号に適用する工程とから成る、請求項11記載の変換方法。

35. 前記の各帯域のスペクトルエンベロープを変換および適用する工程が、低効果サンプリングレートにて再サンプリング信号 $SD(t)$ を作成できるよう、前記帯域での信号を再サンプリングする工程と、

前記帯域での信号の時間スケール処理を行う工程と、

$SD(T)$ の低階位スペクトル分析を行う工程と、

前記の結果たるフィルターを前記の目標励起信号に適用する工程とから成る、請求項11記載の変換方法。

36. さらに、低周波数部分を間引きする工程と、

前記の低周波数部分を分析して、反射係数 K_i を作成する工程と、

前記の音声源音声信号がサンプリングされるレートと同じレートで前記の励起信号をサンプリングする工程と、

補間格子フィルターを使って前記のサンプリングされた励起信号をフィルター処理する工程と、

前記の補間格子フィルターのスペクトル像を除去できるよう、ローパスフィル

ターにて前記の励起信号を後フィルター処理する工程と、

ゲイン値補正を行う工程とから成る、請求項33記載の変換方法。

37. さらに、前記の低周波数部分を間引きする工程と、

前記の低周波数部分を分析して反射係数 k_i を作成する工程と、

前記の低周波数部分の間引き率に合致したレートで前記の励起信号をサンプリングする工程と、

ゲイン値補正を行う工程とから成る、請求項33記載の変換方法。

38. さらに、前記の音声信号を複数の周波数帯域に分割して、前記の帯域に応じてスペクトルエンベロープを個別に変換する工程とから成る、請求項14または28記載の変換方法。

39. さらに、少なくとも50ミリ秒の期間における前記の音声源個人の音声信号のピッチ平均値を算定する工程とから成る、請求項5記載の変換方法。

40. さらに、前記の目標個人の声から励起信号成分を抽出する工程とから成り、前記の励起信号成分の無声部分を音声データで置き換える、請求項一記載の変換方法。

41. さらに、前記の励起信号のピッチ輪郭線を算定する工程とから成る、請求項40記載の変換方法。

42. さらに、前記の励起信号成分を分析セグメントに分割する工程と、前記の分析セグメントのそれぞれが発声信号であるか無声信号であるかを、セグメントのパラメータを、平均セグメント出力値、平均低帯域セグメント出力値、セグメントあたりのゼロ交差数のパラメータグループから選択された閾値と比較することにより判断する工程とから成る、請求項40記載の変換方法。

43. 前記の無声部分を音声データで置き換える工程が、前記無声部分に隣接した発声信号部分の端部間のモーフィング処理をするため正弦波合成を使う工程から成る、請求項40記載の変換方法。

44. 目標たる個人の声から励起信号成分を抽出する方法であって、前記の励起信号の無声部分を音声データで置き換える抽出方法。

45. さらに、前記の励起信号の不適切な発声部分を音声信号で置き換える工程とから成る、請求項44記載の変換方法。

46. 前記の音声データが、下記の1つから由来するものである、請求項44または45記載の変換方法。

(a) 隣接発声部分

(b) 前記の励起信号成分の適切な発声部分

(c) 音声を示すデータのコードブック

47. 前記の置き換える工程が、隣接発声部分からの音声データを補間する工程から成る、請求項44または45記載の変換方法。

48. さらに、それらパラメータは無声部分のピッチ輪郭線や位置情報から成るグループから選択されるものであるような、前記の励起信号成分を特徴付けるパラメータを保存し、前記の音声データで置き換える工程を実行するさいに、前記のパラメータを使う工程とから成る、請求項44、45、46、または、47記載の変換方法。

49. さらに、エンベロープセグメントが不適切な発声信号を示すものであるかどうかを判断する工程とから成る、請求項22記載の変換方法。

50. さらに、不適切な発声セグメントを音声データで置き換える工程とから成る、請求項49記載の変換方法。

51. 前記の音声データが、下記の1つから由来するものである、請求項24または45記載の変換方法。

(a) 隣接発声部分

(b) 前記の励起信号成分の適切な発声部分

(c) 音声を示すデータのコードブック

52. 前記の置き換える工程が、隣接発声部分からの音声データを補間する工程から成る、請求項24または50記載の変換方法。

53. さらに、それらパラメータは無声部分のピッチ輪郭線や位置情報から成るグループから選択されるものであるような、前記の励起信号成分を特徴付けるパラメータを保存し、前記の音声データで置き換える工程を実行するさいに、前記のパラメータを使う工程とから成る、請求項24、50、51、または、52記載の変換方法。

54. 前記のエンベロープセグメントが不適切な発声部分であるかどうかを判断

する工程が、下記の少なくとも1つの工程から成る、請求項49記載の変換方法。

(a) 前記セグメントのLPCゲイン値の振幅を算定する工程

(b) 前記のセグメント内における非常に低い相対エネルギーをもつ高調波の存在を特定する工程

55. さらに、前記の励起信号の不適切な発声セグメントを音声データで置き換える工程とから成る、請求項40記載の変換方法。

56. 前記の音声データが、下記の1つから由来するものである、請求項40または55記載の変換方法。

(a) 隣接発声部分

(b) 前記の励起信号成分の適切な発声部分

(c) 音声を示すデータのコードブック

57. 前記の置き換える工程が、隣接発声部分からの音声データを補間する工程から成る、請求項40または55記載の変換方法。

58. さらに、前記の励起信号のピッチ輪郭線を算定する工程とから成る、請求項55記載の変換方法。

59. さらに、不適切な発声信号セグメントを特定し、それらセグメントを音声データで置き換える工程とから成る、請求項17記載の変換方法。

60. 前記の音声データが、下記の1つから由来するものである、請求項17または59記載の変換方法。

(a) 隣接発声部分

(b) 前記の励起信号成分の適切な発声部分

(c) 音声を示すデータのコードブック

61. 前記の置き換える工程が、隣接発声部分からの音声データを補間する工程から成る、請求項17または59記載の変換方法。

62. 前記の置き換える工程がリアルタイムで行われる、請求項17、24、40、44、または、45記載の変換方法。

63. 前記の置き換える工程が、PSOLA（ピッチ同期重複加算）方法またはレント法を使って前記の音声データのピッチをシフトする工程から成る、請求項

17、24、44、45、50、または、59記載の変換方法。

【発明の詳細な説明】**特定の声を目標とする音声変換****発明の技術分野**

この発明は目標とする声に従って人の声を変換することに関連するものである。もっと具体的には、この発明は目標となる声の録音された情報が変換プロセスをガイドするために使うことのできる変換システムに関連した発明である。さらには歌う人の声を変換しピッチ(音程)やその他の韻律の要素など、目標となる歌い手の声の特質を取り入れて変換することに関連するものである。

背景技術

人の声(ソース音声信号)を別な人の声(ターゲット音声信号)に変換することができれば好ましいと考えられるアプリケーションは数多くある。

この発明はそのような変換をするもので、目標となる声の録音を変換プロセスで利用できるようにアプリケーションに適している。そのようなアプリケーションは自動対話交換(Automatic Dialogue Replacement=ADR)とカラオケがある。カラオケのシステムには正確なピッチ処理が別途必要となるが、話し言葉のシステムには同じ原理が使われるのでここでは説明のためにカラオケのアプリケーションを選んだ。

カラオケは他のアーティストによってポピュラーになった歌をカラオケに参加する人が歌って楽しむことができる。カラオケ用に作られた歌曲は、歌声の部分を取り除き、伴奏の部分だけを残してある。日本ではカラオケは外食産業の次に大きなレジャー産業である。

しかしながら人によっては正しいピッチで歌うことができないためカラオケに参加できない場合がある。

カラオケの遊び方の一つとして歌い手はレコーディングをしたアーティストのスタイルや声を真似たりする。声を変換するという願望はカラオケに限られるものではなく、物まね芸人が例えばエルビスプレスリーの曲を歌うときに重要である。

これまでの音声変換の研究のほとんどは歌声を対象とするものではなく人の話

し言葉に関するものであった。H. KuwabaraとY. Sagisakaの1995年Speech Communication第16巻、スピーカーの音響特性 (Acoustic characteristics of speaker individuality) : 制御と変換 (Control and conversion) は声の特性に関する要素を二つのカテゴリーに分類した。

＊生理学上の要素(即ち声の束の長さ、声門のパルス形状、およびフォルマント (formant) の位置とバンド幅)

＊社会言語学上及び心理学上の要素、及び韻律論の要素(即ち、ピッチ輪郭、言葉の持続時間、タイミングとリズム)

音声変換の研究の大部分は生理学的な要素の直接変換、特に声の広がりや長さの補償 (Vocal tract length compensation)、フォルマントの位置、バンド幅の変換に的が絞られていた。声の特質にとってもっとも重要なのは韻律論的な要素であると認識されるのであるが、現在のスピーチ技術は有効な韻律論的な特性の抽出と取り扱いができていないし、かわりに音声特性を直接的にマッピングすることに焦点を充てていた。

本発明者は変換された声のキャラクターを特定のターゲットの声に類似させる重要なパラメータがターゲット歌手手に依存することを発見した。

歌手によっては、ノート (音符) の始まりでのピッチの輪郭(例えばエルビスプレスリーのすくい上げるような歌い方)はきわめて重要である。他の歌手ではむしろ唸るような声を特徴としている (ルイアームストロングの例)。声の特性を表す重要要素はほかにはビブラートのスタイルがある。

これらの特徴はすべて韻律論的な要素にキーを持っている。生理学上の要素も重要であろう。しかし、我々は生理学上のパラメーターの変換が説得のある音声変換を達成するのには必要とされていないことを発見した。例えば各自のフォルマントの位置やバンド幅を変換することがなくても、聞こえる声の広がりや長さを変換するだけで充分であるかもしれない。

発明の概要

本件の発明は入力される歌手(ソースシンガー)の声のキャラクターを目標される歌手(ターゲットシンガー)のものに変換するための方法や装置を提供するもの

である。この発明はソースシンガーの信号を励起成分（エクサイテーション）と音声の共振成分に分解することに依存している。また、本発明はソースシンガーの励起信号をターゲットシンガーから抽出された励起信号に置き換えることにも依存している。それに加えて、本発明はソースシンガーの音色（Timbre）をターゲットシンガーの音色に、発声の共鳴モデルを修正することにより、シフトする方法を示している。また、さらにはソースシンガーのピッチ輪郭によりよく追従するためのピッチシフトの方法も提示している。

この発明ではまずにターゲット音声信号の励起成分とピッチ輪郭データが必要とされる。これらのものが基本的にはターゲット音声から抽出され、保存され、声の変換に使われる。

この発明はターゲットシンガーのピッチに合わせるためのピッチ修正をするかしないかに関係なく応用できるものである。音声変換のプロセスにピッチ修正も行う場合、ソースシンガーの音声アナログからデジタルデータに変換されそしてセグメントに分けられる。

各セグメントに対して、信号が発声データ、或いは無声データであるかを識別する発声検出器（ボイスンク・ディテクター）が使われる。信号が無声データである場合、その信号がD/A（デジタル・アナログ）コンバーターに送られスピーカーで再生される。発声データがセグメントにある場合には信号がスペクトル・エンベロープ（Spectral Envelope）の形状を決めるために解析される。得られたスペクトル・エンベロープが時間変動合成（Time-varying Synthesis）フィルターを生成するために用いられる。

もし音質シフト（Timbre Shifting）もしくはジェンダーシフト（Gender Shifting）、またはその他の変換も必要な場合、若しくはそうすれば音声変換の結果が改善されるような場合（例えばソースとターゲットの声のスペクトル形状が非常に異なる場合）は、スペクトル・エンベロープが修正されてから、時間変動合成（Time-varying Synthesis）フィルターの生成に使われる。その合成フィルターにターゲットの励起信号を通すことによって変換された音声信号が生成される。

最後に、変換後の音声信号がもとのソース音声信号の振幅エンベロープによって整形される。

ピッチ修正が行わない音声変換のプロセスにおいては、二つの追加ステップが実行される。最初にソース音声のピッチが抽出される。そして、ターゲット励起信号のピッチがソース音声のピッチに追従するようにピッチシフトのアルゴリズムによってシフトされる。

本発明に関して、その他の関連事項を含み、次節の関連手法と適用手法に関する詳細説明の中、および特許請求の範囲の節にてより詳しく記述されている。

図面の簡単な説明

本発明において適用された手法に対する理解は用意されている図と以下の説明を参照することでもっと容易になるであろう。

図1：ターゲット励起信号を生成するプロセッサのブロックダイアグラム。

図2：増強されたターゲット励起信号を生成するプロセッサのブロックダイアグラム。

図3：ピッチ修正を行う音声変換器のブロックダイアグラム。

図4：ピッチ修正なしの音声変換器のブロックダイアグラム（つまりピッチはソース歌手によってコントロールする）。

図5：コンフォーマル・マッピング(Conformal Mapping)によるスペクトルエンベロープの修正。

図6：異なるピッチを持つ音声のスペクトル・エンベロープに表れる違い。

図7：スペクトル・エンベロープの低周波成分と高周波成分に対してそれぞれの修正を図示するブロックダイアグラム。

図8：高いサンプリングレートを持つ信号に対して音声周波数帯域の部分のみを処理するブロックダイアグラム。

最適モードと好適実施例の詳細な説明

図1のブロック図において、目標たる音声信号は、まずデジタルデータに変換される。この工程は、入力信号がすでにデジタル形成である場合には、当然ながら必要ではない。

第1工程は、目標たる音声信号をスペクトル分析するものである。そのスペクトルエンベロープは、目標音声信号のスペクトルエンベロープを平坦化するための時間変動フィルターが作成できるよう、決定され使用される。スペクトル分析を実行する方法は、スペクトルモデルを生成するための従来技術の様々な方法を利用できる。それらスペクトル分析方法には、線形予知法（例として、P. Strobachの「Linear Prediction Theory」、Springer-Verlag、1990年刊を参照）や、応用フィルター処理法（J. I. MakhoulとL. K. Correllの「Adaptive Lattice Analysis of Speech」、IEEE Trans. Acoustics, Speech, Signal Processing, Vol. 29, pp. 654-659、1981年6月刊を参照）などの全極モデル法、ステイグリッツ-マクブライド演算式（K. SteiglitzとL. McBrideの「A Technique for the identification of linear systems」、IEEE Trans. Automatic Control, vol. AC-10, pp. 461-464、1965年刊を参照）などの極ゼロ点モデル法、多帯域励起法を含む変換に基づく方法（D. GriffinとJ. Limの「Multiband excitation vocoder」、IEEE Trans. Acoustics, Speech, Signal Process, vol. 36, pp. 1223-1235、1988年8月刊）、ケプストルに基づく方法（A. OppenheimとR. Schafferの「Homomorphic analysis of speech」、IEEE Trans. Audio Electroacoustics, vol. 16, 1968年6月刊）があげられる。一般的に、全極モデルや極ゼロ点モデルは、格子式や直接式のデジタルフィルターを作成するのに使われている。デジタルフィルターの周波数スペクトルの振幅は、分析から得られるスペクトルエンベロープの振幅値と一致するように選択される。

本好適実施例では、演算の簡易性や安定性から線形予知の自動相関方法を利用する。最初に、目標たる音声信号を、分析セグメントに分割する。自動相関法では、 P 個の反射係数 k_i を作成する。それら反射係数は、全極合成デジタル格子フィルターあるいは全ゼロ分析デジタル格子フィルターのいずれかで、直接の使用が可能である。なお、スペクトル分析 P の級位は、J. MarkelとA. H. Gray Jrの「Linear Prediction of Speech」、Springer-Verlag社、1976年刊に記載されているような、サンプリングレートやその他のパラメータによって決められる。

全極法の直接式実践の応用例は、下記のような時間ドメイン微分関数で表せる。

$$(1) \quad y(k) = x(k) - \sum_{i=1}^P a(i)y(k-i)$$

ただし、 $y(k)$ は現時点でのフィルター出力サンプリング値、 $x(k)$ は現時点での入力サンプリング値、 $a(i)$ は直接式フィルター係数である。それら係数は、反射係数 k_i の値から算出する。その全極合成のための対応する z ドメイン変換関数は、

$$(2) \quad H(z) = \frac{1}{1 + \sum_{i=1}^P a(i)z^{-i}}$$

となる。

補完全ゼロ点分析フィルターには、下記のような微分関数が備わっている。

$$(3) \quad y(k) = x(k) - \sum_{i=1}^P a(i)x(k-i)$$

および、 z ドメイン変換関数は、次のとおりである。

$$(4) \quad H(z) = 1 + \sum_{i=1}^P a(i)z^{-i}$$

直接式格子フィルターであろうが、その他のデジタルフィルターの実行であろうが、目標音声信号は、音声変換事例に適した平坦スペクトルをもつ励起信号を算定するため、分析フィルターで処理される。音声変換器で使うために、その励起信号をリアルタイム演算、または、予め演算しておいてその後で利用できるような保存することも可能である。目標から由来する励起信号は、目標たる歌手の個性を再生するのに必要な情報だけを保存するような、圧縮形態で保存しても構わない。

音声変換器への改良点として、システムが音声源歌手が生み出したタイミング

エラーをいっそう許容できるようにするため、目標励起信号をさらに処理することも可能である。例えば、音声源歌手が所定の歌を歌うとき、そのフレーズが

その歌の目標たる歌手のフレーズとわずかに異なる場合がある。音声源歌手が、その歌の録音での目標たる歌手の出だしよりも少し早く歌い始めた場合、目標歌手の歌い出しの時点まで、出力を作成するための励起信号が生成されない。音声源歌手は、システムが反応しないことに気づき、その反応遅れに不安をもつようになる。たとえ、歌詞の整合が正確であっても、音声源歌手による無声セグメントが、目標歌手の無声セグメントと正確に一致することはあり得ない。その場合、目標歌手の信号の無声部分からの励起が、出力における発声セグメントを生成するのに使われると、出力の音声が非常に不自然となる。この改良信号処理のゴールは、歌中の各語の前後の無声域内まで励起信号を伸長することであり、歌詞の語の無声域を特定してセグメントのための発声域励起を与えることである。また、変換処理のためには適切でない発声域も存在する。例えば、鼻音には、非常にわずかなエネルギーしかもたない周波数スペクトル内の区域がある。無声域中に発声励起信号を供与する処理は、システムにタイミングエラーに対してより許容度をもたせるため、その不適切な音声域までも含めるよう拡張される。

前記の改良された励起処理システムを、図2に図示する。目標励起信号は、発声セグメントと無声セグメントに分類される複数のセグメントに分割する。本好適実施例では、発声の検知は、平均セグメント出力値、平均低帯域セグメント出力値、セグメントごとのゼロ交差値などのパラメータの検査により実行できる。

1個のセグメントの総平均出力値が、平均出力値の最近最大値の60dBより下の場合、そのセグメントは無声域であると判断される。ゼロ交差の数が8/msより多い場合には、そのセグメントは無声域であると判断される。ゼロ交差の数が5/msより少ない場合には、そのセグメントは発声域であると判断する。最後に、総帯域平均出力値に対する低域平均出力値の比率が0.25より低いと、そのセグメントは無声域であると判定される。それ以外は、発声域であると判断する。

発声検知器は、発声が適切でないような区域（例えば、鼻音）を検出する能力

をもてるよう改良することができる。鼻音を検知する方法には、LPCゲイン値に基づいた方法がある（鼻音は、大きなLPCゲイン値をもつ傾向がある）。不適切な発声域を検出する一般的な方法では、非常に小さな相対エネルギーをもつ高調波を求めることが基本となる。

発声セグメントでは、ピッチが抽出される。無声つまり無音のセグメントや不適切な発声セグメントは、適切な発声区域（例えば、その前後の発声区域）、あるいは、適切な発声音を示すデータのコードブックからの置換発声データで埋める。コードブックは、1つまたはそれ以上の数の目標信号から直接由来する、あるいは、例えばパラメトリックモデルからの間接的な1組のデータから成る。

発声データによる置換が実行できるような方法は、いくつもある。いずれの場合も、そのゴールは、意味ある方法で制限ピッチ輪郭線と合致したピッチ輪郭線をもつような（例えば、歌う場合、置換された音符は伴奏と調和する必要がある）音声信号を作成することである。適用例によっては、補間されたピッチ輪郭線を、方形スプライン補間などを使って自動的に演算することも可能である。本好適実施例では、最初にピッチ輪郭線をスプライン補間により算出して、その後で不満足と思われる部分だけを操作者が手動で固定している。

適切なピッチ輪郭線が得られたなら、次に、無声域や不適切な発声域の除去により残された波形上の隙間を補間ピッチ値で埋める必要がある。それを行う方法も、いくつかある。一例として、適切な発声セグメントからのサンプルをその隙間に転写して、その後に、補間ピッチ輪郭線を使ってピッチシフトを行う。そのようなピッチシフト方法の例としては、例えば、PSOLA（ピッチ同期オーバーラップ追加法）、レント法（Lentの「An Efficient Method for Pitch Shifting Digitally Sampled Sounds」、Computer Music Journal、Vol. 13、No. 4、1989年冬号やGibsonらの方法）、Gibsonらの米国特許第5、231、671号に記載の改善方法などのフォルマント補正ピッチシフト法がある。

ここで強調したいのは、無声域と不適切な発声域のための置換に使う方法が何であっても、候補となる波形部分は目標信号内の適当な場所から得られるという

ことである。例えば、置換処理中に使う候補波形部分つまりセグメントを保存するのに、コードブックを利用することもできる。置換が必要な場合、周辺データへの良好な整合を可能にするセグメントを見つけるのにそのコードブックを調べ、その後、それらセグメントを補間目標ピッチとなるようピッチシフト処理する。さらにまた、無声あるいは適切な発声のない区域の置換は、目標音声信号において

て直接にリアルタイムで行うことが可能なものも注意してほしい。

本好適実施例においては、隙間の両側の波形上でのモーフ処理を行うため、正弦波合成を利用している。正弦波合成は、スピーチ圧縮などの分野で広く使われてきた（例えば、D.W.GriffinとJ.S.Limの「Multiband excitation vocoder」、IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36、pp. 1223-1235、1988年8月刊を参照）。スピーチ圧縮においての正弦波合成は、信号セグメントを示すのに必要なビット数を削減するために使われる。その事例では、1つのセグメントのピッチ輪郭線は、一般的に2次または3次補間法を使って補間される。しかしながら、我々の適用例では、圧縮がゴールではなく、（操作者により手動で作成された）予め特定したピッチ輪郭線を追従して、1つの音を別の音へモーフィング処理することであって、それゆえ、本好適実施例では下記に説明するような新規の技法を開発した（ただし、演算式は簡潔化のため連続時間ドメインで示す）。

ここで、時間 t_1 と t_2 の間隔を、正弦波補間で埋めるものとする。まず最初に、ピッチ輪郭線 $w(n)$ を、（自動的に、あるいは、操作者による手動で）決める。そして、ピーク選別の高速フーリエ変換（FFT）を使ったスペクトル分析（例えば、R.J.McAulayとT.F.Quatieriの「Sinusoidal Coding」、Speech Coding, and Synthesis, Elsevier Science B.V., 1995年刊）を t_1 と t_2 で行い、スペクトル振幅値 $A_k(t_1)$ と $A_k(t_2)$ と位相値 $\phi_k(t_1)$ と $\phi_k(t_2)$ を算定するが、ただし、添字の k は高調波の数である。合成信号セグメント $y(t)$ は、下記の式から算出できる。

$$(5) \quad x(t) = \sum_{k=1}^K A_k(t) \cos[\theta_k(t)]$$

ただし、 k は、(セグメントの最長ピッチ期間のサンプル数の長さの半分に設定された)セグメント内の高調波の数である。 $t_1 \leq t \leq t_2$ で時間変動位相を使った我々のモデルは、以下に示す。

$$(6) \quad \theta_k(t) = \theta_k(t_1) + k \int_{t_1}^t [w(t) + r_k(t)] dt + d_k$$

ただし、 $r_k(t)$ は、高調波位相間の相関を削減、ゆえに、感知バズ成分を低減するために使うランダムピッチ成分であって、 d_k は、合成セグメントの開始端と終端での位相を整合させるのに使う線形ピッチ補正項である。セグメント境界部での非連続位相を回避するために $\theta_k(t_1) = \phi(t_1)$ と $\theta_k(t_2) = \phi(t_2)$ が必要であるという事実を使えば、その制約条件を満足させる d_k の最小可能値は、下記のように示すことができる。

$$(7) \quad d_k = \left[v_k T - 2\pi \left\lfloor \frac{2\pi v_k T + \pi}{2\pi} \right\rfloor \right] \frac{1}{T}$$

ただし、 $T = (t_2 - t_1)$ 、および

$$(8) \quad v_k = \left[\phi_k(t_2) - \phi_k(t_1) - \int_{t_1}^{t_2} (w(t) + r_k(t)) dt \right] \frac{1}{T}$$

とする。

前記のランダムピッチ成分 $r_k(t)$ は、合成される隙間に隣接する信号セグメントの予測位相と測定位相との差を算定することにより各高調波で決められた分散値をもつランダム変数のサンプリング、および、その値に比例した分散値の設定から得られる。

最後に、最初に説明した非改良励起抽出と同様に、目標たる励起信号の振幅エンベロープを自動ゲイン値補正を使って平坦化する。

前記の励起信号は、複数の目標音声信号から作成した合成信号でも構わない。この方法では、励起信号には、和音、デュエット、または、伴奏の部分も含めることができる。例えば、男性歌手と女性歌手が同時にデュエットで歌う励起信号は、それぞれが前述のように処理できる。それゆえ、本装置で使う励起信号は、

それら励起信号の和となる。それゆえ、本装置で生成された変換音声信号には、それぞれの目標音声信号から由来する特性（ピッチ、ビブラート、呼吸音など）をもつ各パートから成る両方の発声部分が含まれている。

そして、結果たる基本つまり改良された目標励起信号とピッチデータは、通常は、後で利用するため、音声変換器内に保存される。別の例として、未処理の目

標励起信号を保存しておき、必要なときに、目標励起信号を作成することも可能である。励起の改良を完全に規則に基づいて行うか、あるいは、無音や無声のセグメント中に励起信号を生成するためのピッチ輪郭線やその他の制御値を未処理の目標音声信号と共に保存しておくこともできる。

次に、図3のブロック図を説明する。

音声源の音声信号サンプルがブロック別に、発声か無声かを判断するため分析される。そのブロックに含まれるサンプルの数は、一般的には、ほぼ20ミリ秒の時間間隔に相当するものであって、サンプリングレートが40kHzの場合、20msのブロックには800個のサンプルが含まれる。この分析処理は、時間変動スペクトルエンベロープの現時点での推定値を得るため、周期あるいはピッチ同期を基準にして繰り返す。その繰り返し期間は、サンプルのブロックの時間伸長の期間よりも少ない時間間隔で構わないが、後続の分析が音声サンプルの重複ブロックを利用できることを意味している。

サンプルのブロックが無声入力を示すと判定されると、そのブロックはさらなる処理を行わずに、出力スピーカに送るためデジタル／アナログ変換器に伝送される。サンプルのブロックが発声入力だと判断されると、スペクトル分析を行って、音声信号の周波数スペクトルのエンベロープの推定値を算定する。

音声変換の処理によっては、スペクトルエンベロープの形状を変更することが、望ましい、あるいは、必要となることもある。例えば、音声源の目標音声信号の性別が異なる場合、目標音声信号の音色により密接に整合できるよう、スペクトルエンベロープをスケール操作することにより音声源の声の音色をシフトすることが望ましい。本好適実施例では、スペクトルエンベロープの変更のための選択部分（図3では「変更スペクトルエンベロープ」と表示）で、スペクトル分析

部から得られたエンベロープの周波数スペクトルが変更される。ここで、5つのスペクトル変更の方法を提案する。

第1の方法は、等角写像を(2)式の z ドメイン変換関数に適用することにより元のスペクトルエンベロープを変更する方法である。等角写像により変換関数が変わられ、その結果、下記のような新規の変換関数ができる。

$$(9) \quad H(z) = \frac{\sum_{n=0}^N b(n)z^{-n}}{1 + \sum_{n=1}^N a(n)z^{-n}}$$

等角写像の適用の結果、図5に示されるような変更スペクトルエンベロープが得られる。等角写像のデジタルフィルタへの応用技術の詳細は、A. Constantinidesの「Spectral transformation for digital filters」、IEEE公報、vol. 117、pp. 1585-1590、1970年8月刊に記述されている。その方法の長所は、変換関数の特異点を算定することが不必要なことである。

第2の方法は、デジタルフィルタの変換関数の特異点(極とゼロ点)を見つけて、それら特異点のいずれか、または、全部の位置を変更して、所望のスペクトルエンベロープをもつ新規のデジタルフィルタを作成できるよう変更された新規の特異点を使う方法である。音声信号変更に適用するこの第2の方法は、従来技術で周知である。

スペクトルエンベロープを変更する第3の方法は、別のスペクトルエンベロープ変更工程の必要性を排除するものであって、スペクトル分析の前に音声信号のブロックの時間伸長を変更する方法である。この結果、スペクトル分析の結果として得られるスペクトルエンベロープが、未変更のスペクトルエンベロープの周波数スケール処理されたものとなる。時間スケール処理と周波数スケール処理との関係は、下記のフーリエ変換式により数学的に説明できる。

$$(10) \quad f(at) \leftrightarrow \frac{1}{|a|} F\left(\frac{j\omega}{a}\right)$$

ただし、等式の左側は時間スケール処理信号であって、等式の右側は結果となる周波数スケール処理されたスペクトルである。例えば、在来の分析ブロックの

長さがサンプル800個の場合(20msの信号を示す)、それらサンプルから880個のサンプルを作成するのに補間方法を利用できる。サンプリングレートが固定であるため、これにより時間周期が長くなるよう(22ms)ブロックが時間スケール処理される。時間伸長部を10パーセント長くすることにより、結果と

してスペクトルエンベロープの特性における周波数が10%削減できる。スペクトルエンベロープを変更するこの第3の方法では、必要な演算量が最小となる。

第4の方法は、その内容が本文で参照例として引用されている、S. Seneffの「System to independently modify excitation and/or spectrum of speech waveform without explicit pitch extraction」、IEEE Trans. Acoustics, Speech, Signal Processing, Vol. 30、1982年8月刊に記載のような信号の周波数変換形状を操作する方法である。

第5の方法は、(高い階位をもつ) デジタルフィルタの変換関数を複数の低位数の部分に分解する方法である。それら低位数部分のいずれも、前述した方法を使って変更することが可能である。

目標歌手と音声源歌手とのピッチ差がかなりの量、例えば、1オクターブである場合、それぞれのスペクトルエンベロープに、特に1kHzより下の低帯域において、顕著な差ができるという問題が発生する。例えば、図6のように、低ピッチ発声域では200Hz付近で低周波数共振が、高ピッチ発声域では400Hz付近での高周波数共振が起こる結果となる。その差は、2つの問題を発生させる。

*変換された音声信号における低周波数出力値の低減。

*出力ピッチの高調波付近の周波数をもたないスペクトルピーク値によるシステムノイズの増幅。

これらの問題は、前述のスペクトルエンベロープ変更方法を使って達成できるようなスペクトルエンベロープの低周波数部分を変更することにより緩和できる。スペクトルエンベロープの低周波数部分は、2番目か4番目の方法を使って直接に変更可能である。

また、目標の音声信号を低周波数成分(例えば、1.5kHz以下あるいは同等)と高周波数成分(例えば、1.5kHz以上)とに分割すれば、1番目と3番目

の方法も同じ目的のため利用できる。そして、両成分に対して、図7に図示のように、別のスペクトル分析を行う。さらに、低周波数分析からのスペクトルエンベロープを、ピッチの差、つまり、スペクトルピーク値の位置での差に従って変更する。例えば、目標たる歌手のピッチが200kHzで、音声源歌手のピッチが400kHzである場合、未変更の音声源スペクトルエンベロープは400Hz

付近でピーク値をもち、200Hz付近でピークがないので、200Hz付近でのゲイン値が小さくなって、結果として、前述した第1番目の問題が発生する。それゆえ、スペクトルピークを400Hzから200Hzへ移動させるよう、低周波数エンベロープを変更したのである。

本好適実施例では、下記の手順でスペクトルエンベロープの低周波数部分を変更している。

1. 音声源の音声信号 $S(t)$ は、ほぼ1.5kHz以下の周波数のみをもつ帯域制限信号 $SL(t)$ を作成できるようローパスフィルター処理する。
2. この帯域制限信号 $SL(t)$ を、低率信号 $SD(t)$ を作成できるようほぼ3kHzで再サンプリングする。
- 低階位スペクトル分析（例えば、 $P=4$ ）を $SD(t)$ に対して行い、直接式フィルター係数 $ap(i)$ を算定する。
3. それら係数を、目標の音声信号のピッチと音声源の音声信号のピッチとの比率に比例してスペクトルをスケール処理するため、等角写像法を使って変更する。
4. 結果となるフィルターを、補間フィルター処理法を使って（元のサンプリングレートをもつ）信号 $SL(t)$ に適用する。

上記の方法を使えば、図7に図示されているように、信号の低周波数と高周波数の部分が個別に処理した後で、合成されて出力信号を作成することができる。図7に示す装置は、低周波数のスペクトルエンベロープ、あるいは、高周波数のスペクトルエンベロープだけを変更するのにも利用できる。そのようにして、高周波数共振の音色に影響することなく低周波数を変更したり、または、高周波数共振の音色だけを変更することが可能となる。また、両方のスペクトルエンベロ

ープを同時に変更することもできる。

上記のスペクトルエンベロープの低周波数域に関する問題を排除するのに利用できる別の方法として、スペクトルピークの帯域幅を増加させる方法がある。これは、以下のような従来技術の方法で達成できるものである。

＊帯域幅の拡張。

＊選択した極半径の変更。

＊フィルター係数を算定する前における自動相関ベクトルのウィンドウ処理。

高忠実度のオーディオ装置は、一般的に、スピーチ分析つまり符号化装置よりも高いサンプリングレートを使っている。スピーチにおいては、支配スペクトル成分の大部分の周波数が10kHzより低いことが、その理由である。高忠実度装置にて高サンプリングレートを使えば、デジタルフィルターにより信号を高周波数信号（例えば、10kHzより大きい）と低周波数信号（例えば、10kHzより小さいか同等）に分割しても、前記のスペクトル分析Pの階位を減らせる。その後、スペクトル分析の前にこの低周波数信号をダウンサンプリングすることにより低サンプリングレートにするので、分析の階位を低くすることができるのである。

低サンプリングレートおよび低分析階位の結果、演算処理量を削減できる。本好適実施例では、入力音声信号が40kHz以上の高いレートでサンプリングされる。そして、図8のように、信号を2つの同じ幅の周波数帯に分割する。低周波数部分は間引きされて、反射係数 k_1 を作成できるよう分析処理される。また励起信号も同じ高レートでサンプリングしてから、補間格子フィルター（つまり、単位遅延を2つの単位遅延で置換できる格子フィルター）にてフィルター処理する。その後、信号を後フィルター処理して補間格子フィルターのスペクトル像を除去して、ゲイン値補正が行われる。その結果の信号は、低周波数成分の変換された音声信号となる。補間フィルター処理法は、再サンプリング中におけるエイリアシングによる歪をより完全に除去できるため、従来のダウンサンプリングフィルター処理ーアップサンプリングという一連の処理法に代えて、採用している。励起信号を間引きレートに合った低レートでサンプリングすれば、補間格子フィ

ルターの必要性はなくなる。本発明においては2つの異なるサンプリングレートを同時に使用するのが好ましく、その結果、必要演算量を低減できる。

そして、ゲイン値補正された高周波数信号と変換された低周波数成分とを合成すれば、最終的な出力信号が得られるのである。この方法は、図7に図示した方法と共同で行うこともできる。

このように、スペクトルエンベロープは、上記の各方法およびそれらの組み合わせによって修正変更が可能である。そして、変更されたスペクトルエンベロー

プは、対応した周波数応答性をもつ時間変動合成デジタルフィルタを作成するのに利用される。「スペクトルエンベロープ適用」と表示されたブロック部で、励起信号抽出処理工程の結果として生成された目標励起信号がこのデジタルフィルタに掛けられる。本好適実施例では、このフィルタ処理に格子デジタルフィルタを使っている。フィルタの出力は、離散時間様式の所望の変換音声信号となる。

「振幅エンベロープ適用」と表示された図3のブロック部は、変換された音声信号の振幅を音声源の音の振幅に追従させるものである。このブロック部では、以下のような補助演算処理が必要となる。

- * デジタル化された音声源音声信号 L_s のレベル。
- * デジタル化された目標励起信号 L_e のレベル。
- * スペクトルエンベロープ L_1 の適用後の信号のレベル。

これらレベル値は、合成フィルタを通過させた後で元の信号に適用される出力信号振幅レベル値を算出するのに使われる。

本好適実施例では、下記のような再帰アルゴリズムにより各レベル値が演算される。

* 32個のサンプルの i 番目のフレームのフレームレベル値 $L_f(i)$ は、フレーム内のサンプルの絶対値の最大値として算定する。

* 減衰された先行レベル値は、 $L_d(i) = 0.99L(i-1)$ として算定する。

* レベル値は、 $L(i) = \max \{L_f(i), L_d(i)\}$ として算定する。

現時点での出力フレームに適用すべき振幅エンベロープも、再帰アルゴリズム

で演算する。

*未円滑化の振幅補正值 $A_r(i) = L_s L_e / L_f$ の算定。

*円滑化された振幅補正值 $A_s(i) = 0.9 A_s(i-1) + 0.1 A_r(i)$ の算定。

本アルゴリズムでは、装置の処理動作遅延を補正するために、遅延値 L_s と L_e を使っている。

フレームツーフレーム値 A_s はフレーム間で線形補間処理され、円滑に変動する振幅エンベロープを作成する。スペクトルエンベロープ適用ブロック部からのサンプルは、それぞれこの時間変動エンベロープで積算される。

図4には、音声源音声信号のピッチが保持される例が図示されている。その例では、音声源音声信号のピッチを決定する。それを実行する方法が、本文で内容が参照として引用されている、Gibsonらの米国特許第4、688、464号に開示されている。目標たる励起信号は、変更あるいは未変更の音声源スペクトルエンベロープを励起信号に適用するよりも前に、音声源音声信号のピッチを追従するのに必要な量だけピッチがシフトされる。この目的に適したピッチシフトの方法は、Gibsonらの米国特許第5、567、901号に開示されており、本文で内容が参照として引用されている。ここで、本操作モードにより音声源歌手がその出力をより良く制御できる一方で、目標歌手の個性がビブラートやピッチスクーピングなどの素早いピッチ変化を行うような特徴をもつ場合には、変換処理の効果性が著しく低減する恐れがある。その特徴的な急激なピッチ変化によるロスを防止するため、ピッチ検出処理で、ピッチシフト量を算定するさいに長期平均値を使うことも可能である。ピッチデータを、目標歌手の特性に従って50msから500msの範囲での平均化する。平均値演算は、新規の音符が検知されるたびにリセットする。場合によっては、キー変動を行うため目標励起ピッチを固定量だけシフトして、音声源歌手のピッチを無視することもできる。

本発明の範囲を逸脱することなく本好適実施例のその他の変更例が実現できることは、当業者にとっては明白であろう。また、本発明のアプローチが、歌唱での音声に限定されるものではなく、スピーチにも同様に適用可能であることも明らかであろう。

【図1】

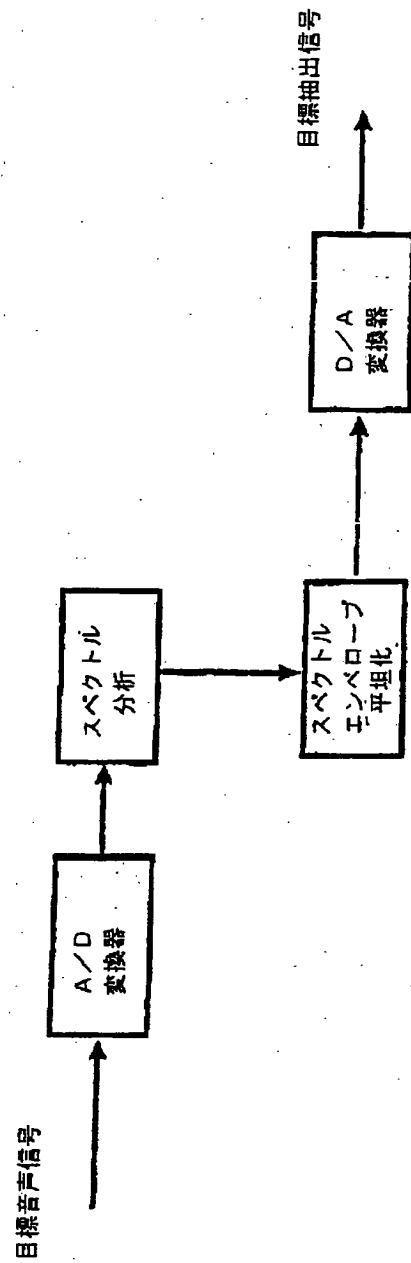


図1

【図2】

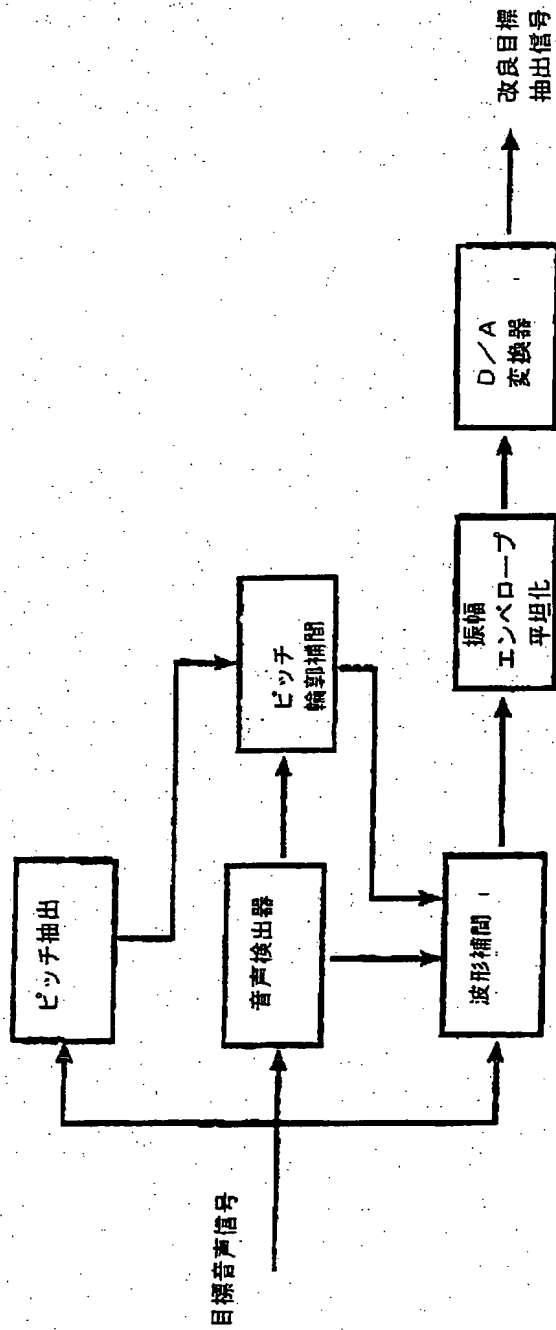


図2

【図3】

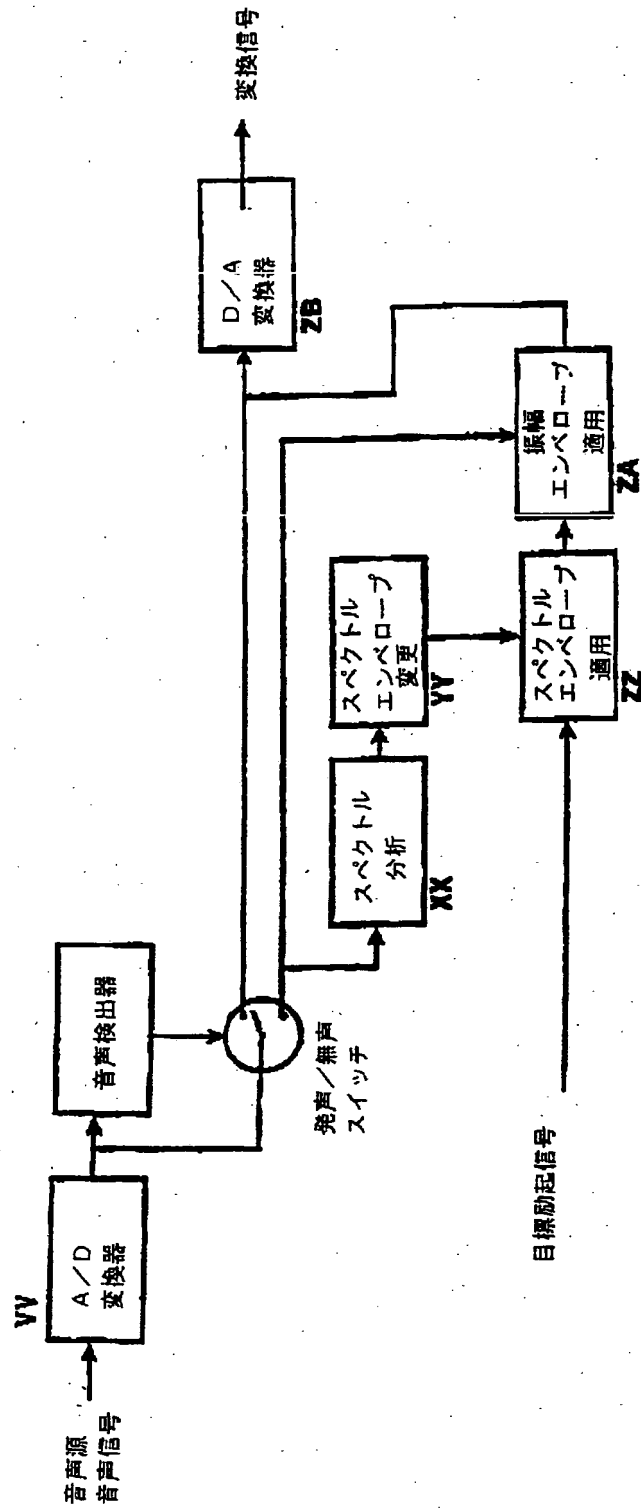


図3

【図4】

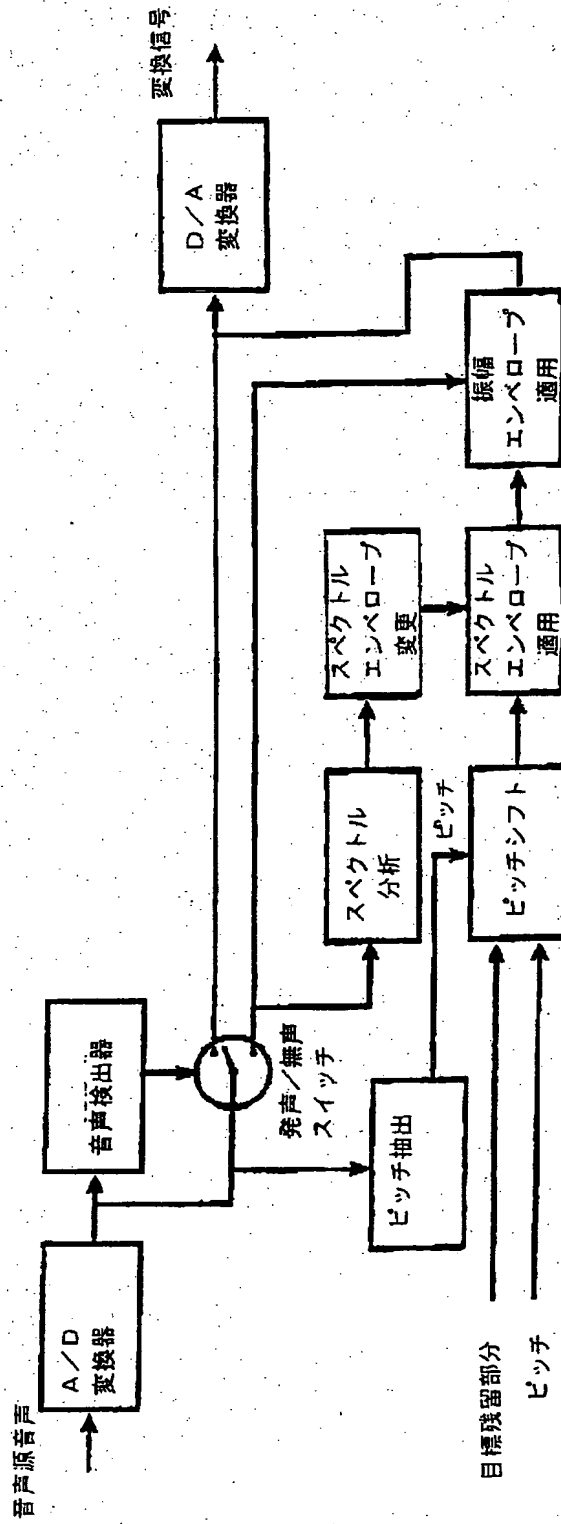


図4

【図5】

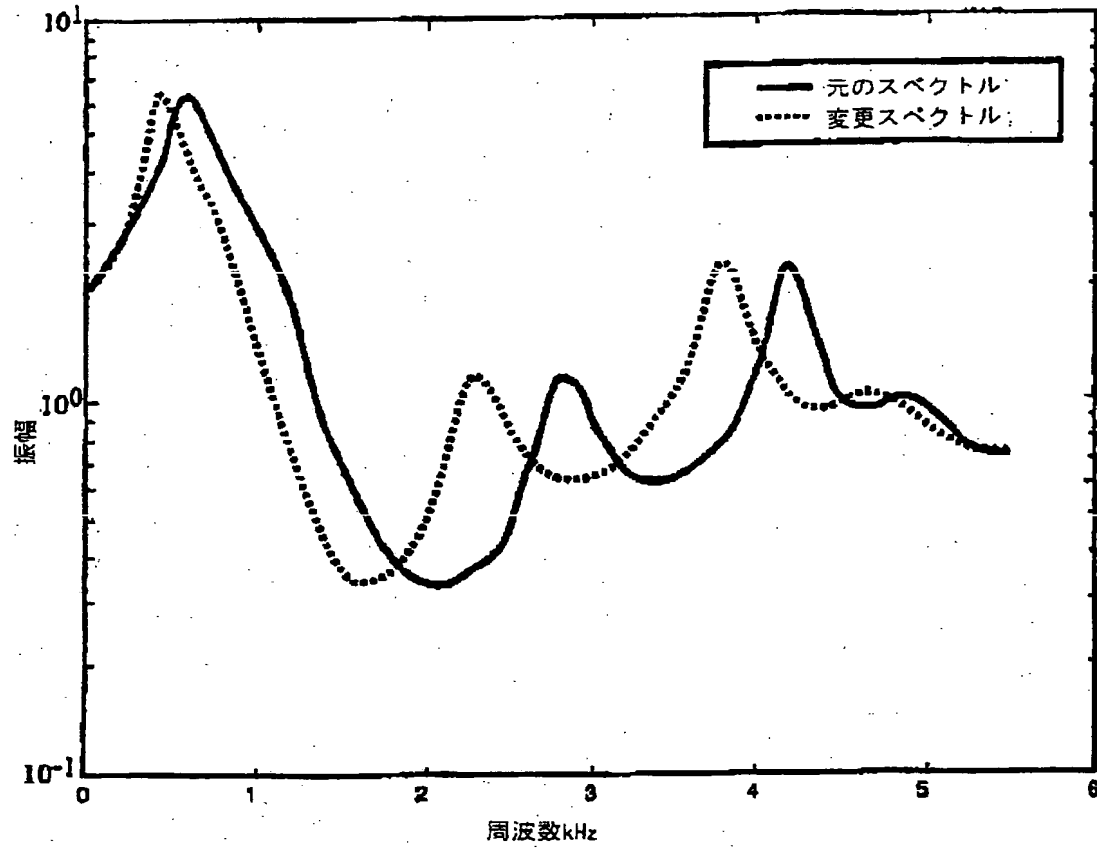


図5

【図6】

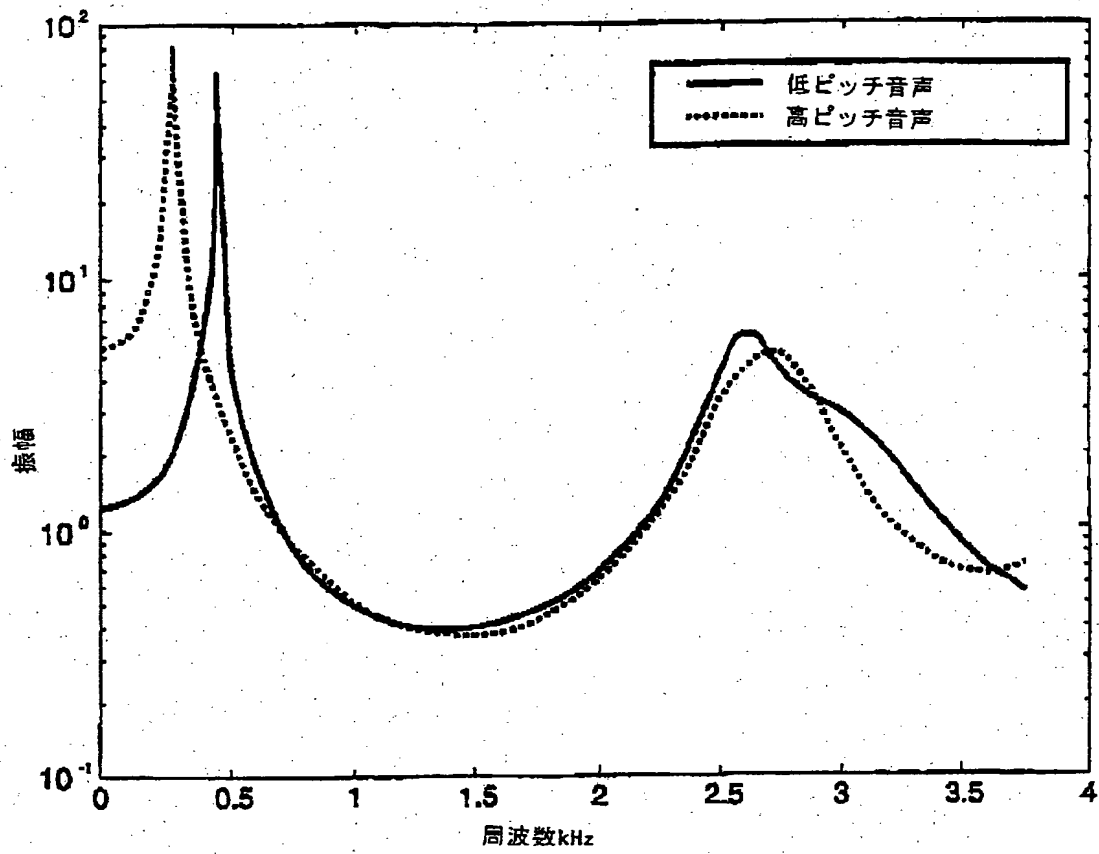


図6

【図7】

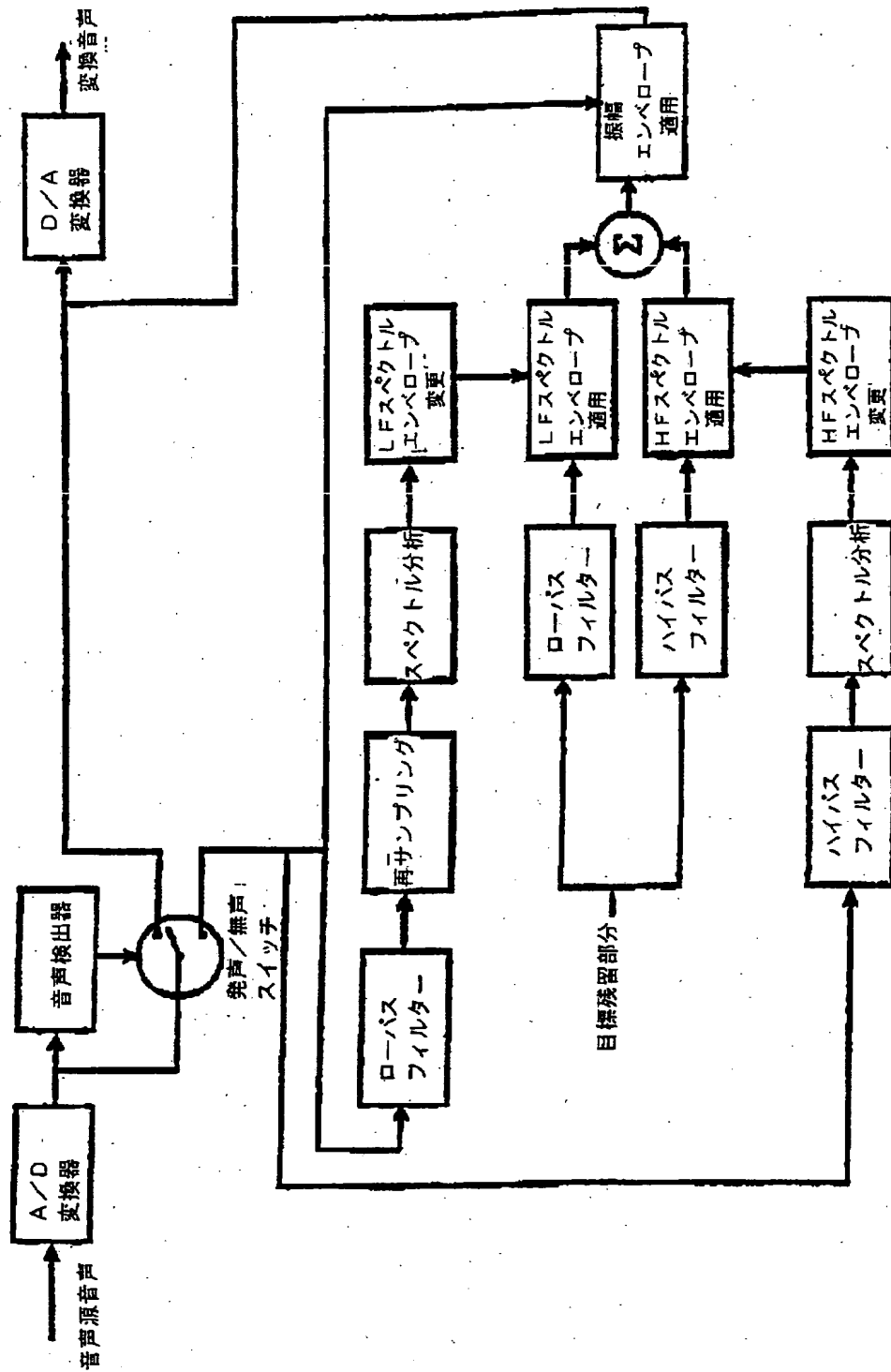


図7

【図8】

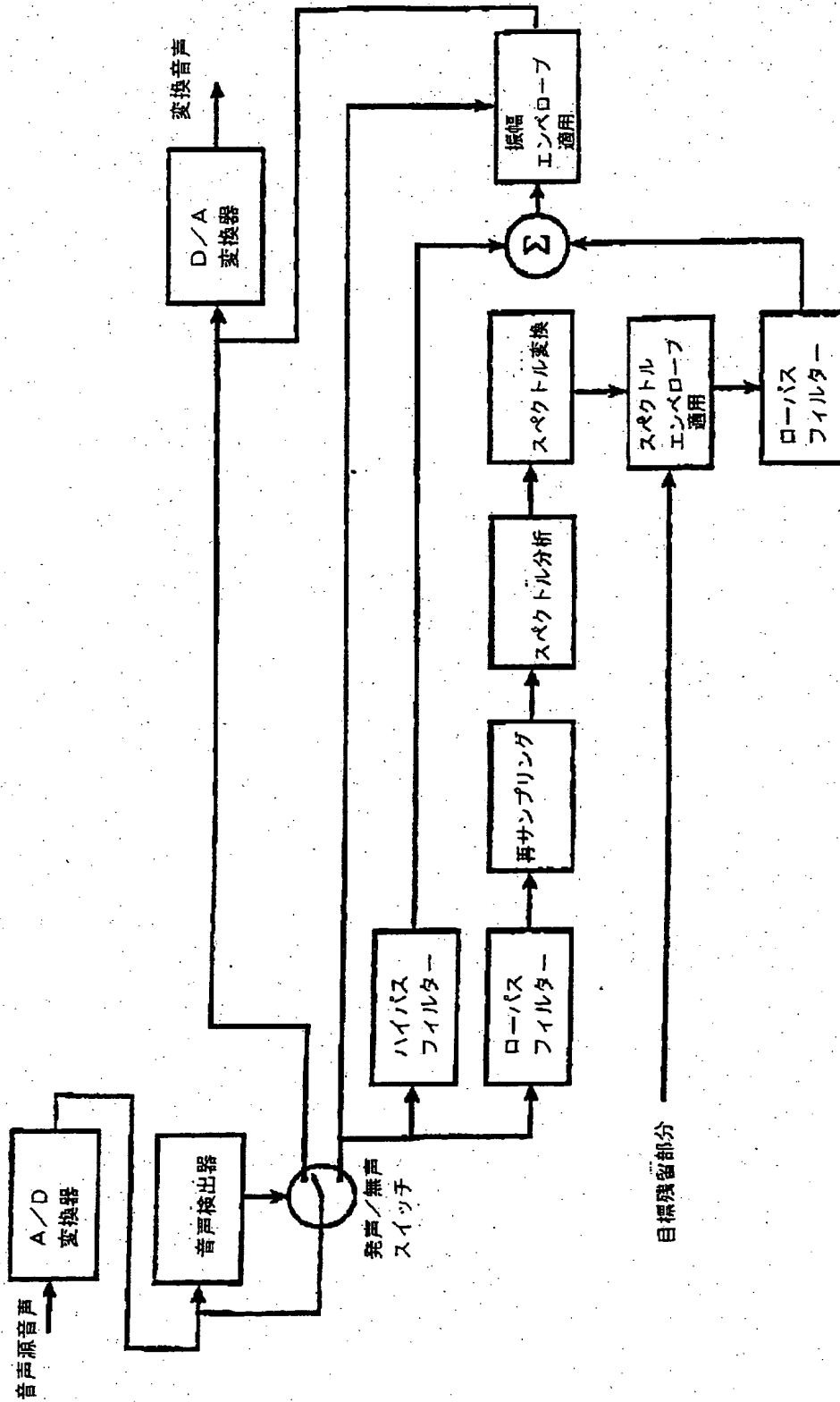


図 8

【国際調査報告】

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 G10H1/36 G10L3/02		Int. Appl. No. PCT/CA 98/00406
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols): IPC 6 G10H G10L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim no.
P.A	PATENT ABSTRACTS OF JAPAN vol. 097, no. 011, 28 November 1997 & JP 09 198091 A (YAMAHA CORP), 31 July 1997, see abstract -& US 5 750 912 A (MATSUMOTO) 12 May 1998 see column 1, line 60 - column 3, line 34 see column 7, line 13 - line 55; figure 8 ---	1-63
A	WO 93 18505 A (WALT DISNEY PROD) 16 September 1993 see page 8, line 4 - page 11, line 22; figure 2 ---	1-63
A	WO 90 13887 A (UNIV LELAND STANFORD JUNIOR) 15 November 1990 see page 3, line 3 - page 8, line 24; figures 1,2 ---	1-63
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		<input checked="" type="checkbox"/> Patent family members are listed in annex
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "A" document member of the same patent family		
Date of the actual completion of the international search 30 June 1998		Date of mailing of the international search report 09/07/1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 apo nl. Fax: (+31-70) 340-3016		Authorized officer Pulluud, R

Form PCT/ISA/210 (second sheet) July 1992

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/CA 98/00406

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 536 902 A (SERRA XAVIER ET AL) 16 July 1996 see column 8, line 39 - column 14, line 16; figures 1-4	1-63
A	MIZUNO H ET AL: "VOICE CONVERSION BASED ON PIECEWISE LINEAR CONVERSION RULES OF FORMANT FREQUENCY AND SPECTRUM TILT" 19 April 1994, PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, SIGNAL PROCESSING (ICASSP), SPEECH PROCESSING 1, ADELAIDE, APR. 19 - 22, 1994, VOL. VOL. 1, PAGE(S) I-469 - I-472, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS XP000529420 see page I-469, left-hand column, line 16 - page I-470, right-hand column, line 8; figures 1,2	1-63

Form PCT/ISA/210 (continuation of second sheet) (July 1982)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No.

PCT/CA 98/00405

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9318505 A	16-09-1993	US 5327521 A	05-07-1994
WO 9013887 A	15-11-1990	US 5029509 A	09-07-1991
		AU 5532890 A	29-11-1990
US 5536902 A	16-07-1996	JP 7325583 A	12-12-1995

Form PCT/CA/210 (patent family annex) July 1992

フロントページの続き

(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), AP(GH, GM, KE, LS, MW, SD, SZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW

(72)発明者 ルビニ, ピーター, ロナルド
カナダ国 プリティッシュ コロンビア
ブイ8エル 5エイチ8, ノース スナニ
ッチ, トライオン ロード 2365

(72)発明者 シュバク, デール, ジョン
カナダ国 プリティッシュ コロンビア
ブイ8エヌ 2シー9, ビクトリア, ジャ
マイカ ロード 1445

This Page Blank (uspto)